Amendments to the Claims:

This listing of claims will replace all prior versions, and listings, of claims in the application:

Listing of Claims:

1. (Currently Amended) A computer-implemented method of identifying whether a sequence of terms is a semantic unit, the method comprising:

receiving the sequence of terms in a memory;

calculating a first value representing a coherence of terms in the sequence;

calculating a second value representing variation of context in which the sequence occurs;

determining whether the sequence is a semantic unit based at least in part on the first and second values; and

outputting an indication of whether the sequence is a semantic unit for use in a processor.

2. (Original) The method of claim 1, wherein the coherence of the terms in the sequence is calculated relative to a collection of documents.

3. (Original) The method of claim 2, wherein the coherence of the terms in the sequence is calculated as a likelihood ratio that defines a probability of the sequence occurring in the collection of documents relative to parts of the sequence occurring.

4. (Original) The method of claim 2, wherein the coherence of the terms in the sequence is calculated as:

$$LR(A,B) = \frac{L(f(B),N)}{L(f(AB),f(A)) \cdot L(f(\sim AB),f(\sim A))},$$

where $f(A)$ defines a number of occurrences of term A in the collection of documents, $f(\sim A)$ defines a number of occurrences of a term other than term A in the collection of documents, $f(B)$ defines a number of occurrences of term B in the collection of documents, N defines a total number of events in the collection of documents, $f(AB)$ defines a number of times term A is followed by term B in the collection of documents, and $f(\sim AB)$ is a number of times a term other than A is followed by term B in the collection of documents, wherein

$$L(k,n) = \left(\frac{k}{n}\right)^{k} \cdot \left(1-\frac{k}{n}\right)^{(n-k)}.$$

5. (Original) The method of claim 1, wherein the coherence of the terms in the sequence are defined as not being sufficient unless a threshold is met.

6. (Original) The method of claim 5, wherein the threshold is defined as: $f(AB) > \frac{f(A) \cdot f(B)}{N}$, where $f(A)$ defines a number of occurrences of term A in the collection of documents, $f(B)$ defines a number of occurrences of term B in the collection

of documents, *N* defines a total number of events in the collection of documents, and *f(AB)* defines a number of times term A is followed by term *B* in the collection of documents.

7. (Original) The method of claim 1, wherein the variation of context in which the sequence occurs is calculated relative to a collection of documents.

8. (Original) The method of claim 7, wherein the variation of context in which the sequence occurs is calculated as a measure of entropy of the context of the sequence.

9. (Original) The method of claim 7, wherein the variation of context in which the sequence occurs, *H(S)*, is calculated as

$$HM(S) = MIN(HL(S), HR(S)),$$

$$HLM(S) = -\sum_{w} \frac{f(wS)}{f(S)} \cdot \log\left(\frac{f(wS)}{f(S)}\right),$$

and

$$HR(S) = -\sum_{w} \frac{f(Sw)}{f(S)} \cdot \log\left(\frac{f(Sw)}{f(S)}\right),$$

where *MIN* defines a minimum operation, *S* represents the sequence, *f(wS)* defines a number of times a particular term, *w*, appears in the collection of documents followed by the sequence, *f(Sw)* refers to a number of times the sequence is followed by *w* in the

collection of documents, and *f(S)* refers to a number of times the sequence *S* is present in the collection of documents.

10. (Original) The method of claim 7, wherein the variation of context in which the sequence occurs, *HM(S),* is calculated as

$$HM(S) = MIN(HLM(S), HRM(S)),$$

where MIN defines a minimum operation, HLM*(S)* is defined as a minimum of $1 - \frac{f(wS)}{f(S)}$ for each term w in the collection of documents, *HRM(S)* is defined as a minimum of $1 - \frac{f(Sw)}{f(S)}$ for each term *w* in the collection of documents, *f(wS)* defines a number of times a particular term, *w,* appears in the collection of documents followed by the sequence, *f(Sw)* refers to a number of times the sequence is followed by *w* in the collection of documents, and *f(S)* refers to a number of times the sequence is present in the collection of documents.

11. (Original) The method of claim 7, wherein the variation of context in which the sequence occurs, *HC(S),* is calculated as

$$HC(S) = MIN(HLC(S), HRC(S)),$$

where MIN defines a minimum operation, *HLC(S)* is defined as $\sum_{w} \delta(wS)$ and *HRC(S)* is defined as $\sum_{w} \delta(Sw)$, where $\delta(X)$ is defined as one if sequence *X* occurs in the

collection of documents and zero otherwise, where wS refers to a particular word followed by the sequence, and where Sw refers to the sequence followed by a word.

12. (Original) The method of claim 7, wherein the variation of context in which the sequence occurs, *HP(S),* is calculated as

$$HP(S) = MIN(HLP(S), HRP(S))$$

where MIN defines a minimum operation, *HLP(S)* is defined as the number of continuations to the left of the sequence that cover a predetermined percentage of all cases in the collection of documents and *HRP(S)* is defined as the number of continuations to the right of the sequence that cover the predetermined percentage of all cases in the collection of documents.

13. (Original) The method of claim 1, wherein determining whether the sequence is a semantic unit includes comparing the first and second values to first and second thresholds and identifying the sequence as a semantic unit when the first and second values satisfy the first and second thresholds.

14. (Original) The method of claim 1, wherein the sequence includes three or more words.

15. (Original) The method of claim 1, further including:

applying one or more rules to the sequence, and

wherein determining whether the sequence is a semantic unit is further based at least in part on the application of the one or more rules.

16. (Currently Amended) A device comprising:

a receiving component configured to receive a sequence of terms;

a coherence component configured to calculate a coherence of multiple terms in [[a]] the sequence of terms;

a variation component configured to calculate a variation of context terms in a collection of documents in which the sequence occurs; and

a decision component configured to determine whether the sequence constitutes a semantic unit based at least in part on results of the coherence component and the variation component, and output an indication of whether the sequence constitutes a semantic unit for use in a processor.

17. (Original) The device of claim 16, wherein the context terms include terms to the left and right of the sequence.

18. (Original) The device of claim 16, wherein the coherence of the terms in the sequence is calculated relative to the collection of documents.

19. (Currently amended) The ~~method~~ device of claim 18, wherein the coherence of the terms in the sequence is calculated as a likelihood ratio that defines a

probability of the sequence occurring in the collection of documents relative to parts of the sequence occurring.

20. (Original) The device of claim 16, wherein the variation of context in which the sequence occurs is calculated as a measure of entropy of the context of the sequence.

21. (Original) The device of claim 20, wherein the variation of context in which the sequence occurs, *H(S)*, is calculated as

$$H(S) = MIN\left(HL(S), HR(S)\right),$$

$$HL(S) = -\sum_{w} \frac{f(wS)}{f(S)} \cdot \log\left(\frac{f(wS)}{f(S)}\right),$$

and

$$HR(S) = -\sum_{w} \frac{f(Sw)}{f(S)} \cdot \log\left(\frac{f(Sw)}{f(S)}\right),$$

where *MIN* defines a minimum operation, *S* represents the sequence, *f(wS)* defines a number of times a particular term, *w*, appears in the collection of documents followed by the sequence, *f(Sw)* refers to a number of times the sequence is followed by *w* in the collection of documents, and *f(S)* refers to a number of times the sequence *S* is present in the collection of documents.

22. (Original) The device of claim 20, wherein the variation of context in which the sequence occurs, *HM(S)*, is calculated as

$$HM(S) = MAX(HLM(S), HRM(S)),$$

where MIN defines a minimum operation, HLM*(S)* is defined as a minimum of $1 - \frac{f(wS)}{f(S)}$ for each term w in the collection of documents, *HRM(S)* is defined as a minimum of $1 - \frac{f(Sw)}{f(S)}$ for each term *w* in the collection of documents, *f(wS)* defines a number of times a particular term, *w*, appears in the collection of documents followed by the sequence, *f(Sw)* refers to a number of times the sequence is followed by *w* in the collection of documents, and *f(S)* refers to a number of times the sequence is present in the collection of documents.

23. (Original) The device of claim 20, wherein the variation of context in which the sequence occurs, *HC(S),* is calculated as

$$HC(S) = MIN(HLC(S), HRC(S)),$$

where MIN defines a minimum operation, *HLC(S)* is defined as $\sum_{w} \delta(wS)$ and *HRC(S)* is defined as $\sum_{w} \delta(Sw)$, where $\delta(X)$ is defined as one if sequence *X* occurs in the document collection and zero otherwise, where wS refers to a particular word followed by the sequence, and where Sw refers to the sequence followed by a word.

24. (Original) The device of claim 20, wherein the variation of context in which the sequence occurs, *HP(S),* is calculated as

$$HP(S) = MIN(HLP(S), HRP(S))$$

where MIN defines a minimum operation, *HLP(S)* is defined as the number of continuations to the left of the sequence that cover a predetermined percentage of all cases in the collection of documents and *HRP(S)* is defined as the number of continuations to the right of the sequence that cover the predetermined percentage of all cases in the collection of documents.

25. (Previously presented) The device of claim 16, wherein the decision component is further configured to compare the results of the coherence component and the variation component to threshold values and identify the sequence as a semantic unit based at least in part on the comparisons.

26. (Original) The device of claim 16, further comprising:

a heuristics component configured to apply one or more predefined rules to the sequence, wherein the decision component is further configured to determine whether the sequence constitutes a semantic unit based at least in part on application of the one or more rules.

27. (Original) The device of claim 26, wherein the one or more rules are exclusionary rules that determine when certain sequences are not semantic units.

28. (Currently Amended) A device comprising:

means for receiving a sequence of terms;

means for calculating a first value representing a coherence of terms in [[a]] the sequence of terms;

means for calculating a second value representing variation of context in which the sequence occurs;

means for determining whether the sequence is a semantic unit based at least in part on the first and second values; and

means for outputting an indication of whether the sequence is a semantic unit for use in a processor.

29. (Previously presented) A computer-readable memory device that includes programming instructions configured to control at least one processor, the computer-readable memory device comprising:

instructions for calculating a first value representing a coherence of terms in a sequence of terms;

instructions for calculating a second value representing variation of context in which the sequence occurs;

instructions for determining whether the sequence is a semantic unit based on the first and second values; and

instructions for outputting an indication of whether the sequence is a semantic unit.

30. (New) The computer-readable memory device of claim 29, wherein the coherence of the terms in the sequence is calculated relative to a collection of documents.

31. (New) The computer-readable memory device of claim 30, wherein the coherence of the terms in the sequence is calculated as a likelihood ratio that defines a probability of the sequence occurring in the collection of documents relative to parts of the sequence occurring.

32. (New) The computer-readable memory device of claim 30, wherein the coherence of the terms in the sequence is calculated as:

$$LR(A,B) = \frac{L(f(B),N)}{L(f(AB),f(A)) \cdot L(f(\sim AB),f(\sim A))},$$

where $f(A)$ defines a number of occurrences of term A in the collection of documents, $f(\sim A)$ defines a number of occurrences of a term other than term A in the collection of documents, $f(B)$ defines a number of occurrences of term B in the collection of documents, N defines a total number of events in the collection of documents, $f(AB)$ defines a number of times term A is followed by term B in the collection of documents, and $f(\sim AB)$ is a number of times a term other than A is followed by term B in the collection of documents, wherein

$$L(k,n) = \left(\frac{k}{n}\right)^{k} \cdot \left(1 - \frac{k}{n}\right)^{(n-k)} .$$

33. (New) The computer-readable memory device of claim 29, wherein the coherence of the terms in the sequence are defined as not being sufficient unless a threshold is met.

34. (New) The computer-readable memory device of claim 33, wherein the threshold is defined as: $f(AB) > \frac{f(A) \cdot f(B)}{N}$, where $f(A)$ defines a number of occurrences of term A in the collection of documents, $f(B)$ defines a number of occurrences of term B in the collection of documents, N defines a total number of events in the collection of documents, and $f(AB)$ defines a number of times term A is followed by term B in the collection of documents.

35. (New) The computer-readable memory device of claim 29, wherein the variation of context in which the sequence occurs is calculated relative to a collection of documents.

36. (New) The computer-readable memory device of claim 35, wherein the variation of context in which the sequence occurs is calculated as a measure of entropy of the context of the sequence.

37. (New) The computer-readable memory device of claim 35, wherein the variation of context in which the sequence occurs, *H(S)*, is calculated as

$$HM(S) = MIN(HL(S), HR(S)),$$

$$HLM(S) = -\sum_{w} \frac{f(wS)}{f(S)} \cdot \log\left(\frac{f(wS)}{f(S)}\right),$$

and

$$HR(S) = -\sum_{w} \frac{f(Sw)}{f(S)} \cdot \log\left(\frac{f(Sw)}{f(S)}\right),$$

where *MIN* defines a minimum operation, *S* represents the sequence, *f(wS)* defines a number of times a particular term, *w*, appears in the collection of documents followed by the sequence, *f(Sw)* refers to a number of times the sequence is followed by *w* in the collection of documents, and *f(S)* refers to a number of times the sequence *S* is present in the collection of documents.

38. (New) The computer-readable memory device of claim 35, wherein the variation of context in which the sequence occurs, *HM(S)*, is calculated as

$$HM(S) = MIN(HLM(S), HRM(S)),$$

where MIN defines a minimum operation, HLM*(S)* is defined as a minimum of $1 - \frac{f(wS)}{f(S)}$ for each term w in the collection of documents, *HRM(S)* is defined as a minimum of $1 - \frac{f(Sw)}{f(S)}$ for each term *w* in the collection of documents, *f(wS)* defines a number of times a particular term, *w*, appears in the collection of documents followed by the sequence, *f(Sw)* refers to a number of times the sequence is followed by *w* in the

collection of documents, and *f(S)* refers to a number of times the sequence is present in the collection of documents.

39. (New) The computer-readable memory device of claim 35, wherein the variation of context in which the sequence occurs, *HC(S),* is calculated as

$$HC(S) = MIN(HLC(S), HRC(S)),$$

where MIN defines a minimum operation, *HLC(S)* is defined as $\sum_{w} \delta(wS)$ and *HRC(S)* is defined as $\sum_{w} \delta(Sw)$, where $\delta(X)$ is defined as one if sequence *X* occurs in the collection of documents and zero otherwise, where wS refers to a particular word followed by the sequence, and where Sw refers to the sequence followed by a word.

40. (New) The computer-readable memory device of claim 35, wherein the variation of context in which the sequence occurs, *HP(S),* is calculated as

$$HP(S) = MIN(HLP(S), HRP(S))$$

where MIN defines a minimum operation, *HLP(S)* is defined as the number of continuations to the left of the sequence that cover a predetermined percentage of all cases in the collection of documents and *HRP(S)* is defined as the number of continuations to the right of the sequence that cover the predetermined percentage of all cases in the collection of documents.

41. (New) The computer-readable memory device of claim 29, wherein the instructions for determining whether the sequence is a semantic unit include instructions

for comparing the first and second values to first and second thresholds and identifying the sequence as a semantic unit when the first and second values satisfy the first and second thresholds.

42. (New) The computer-readable memory device of claim 29, wherein the sequence includes three or more words.

43. (New) The computer-readable memory device of claim 29, further including:

instructions for applying one or more rules to the sequence, and

wherein the instructions for determining whether the sequence is a semantic unit are further based at least in part on the application of the one or more rules.